

# 雑音環境下における 隠れマルコフモデルを用いた和音認識に関する検討

植木 淳宏<sup>†</sup>越田 俊介<sup>†</sup>阿部 正英<sup>†</sup>川又 政征<sup>†</sup><sup>†</sup> 東北大学大学院工学研究科電子工学専攻

## 1 まえがき

和音は音楽の三大要素の1つである和声から派生する音楽要素の1つである。音楽音響信号からの和音認識は、自動採譜や音楽情報検索の分野へ応用されている。近年主流となっている和音認識の手法は、和音進行を隠れマルコフモデル (Hidden Markov Model: HMM) でモデル化することにより認識する手法である [1]。HMM を用いた認識システムでは、音響モデルの学習時と認識時の雑音環境に差がある場合、認識性能は大きく劣化することが知られている。そのため、HMM を用いた音声認識の分野では耐雑音性に優れたモデルの研究が盛んに行われている。しかしながら、和音認識の分野において、雑音環境が認識性能に与える影響については未だ十分に研究されていない。

本稿では、白色ガウス雑音を付加した学習データと評価データを用いて評価実験を行うことにより、雑音が認識率に与える影響について検討する。実験の結果、モデルの作成に用いる学習データの SNR が評価データの SNR と近い場合に高い認識率が得られた。

## 2 音楽音響信号からの和音認識

本稿では、Sheh らが提案した和音認識の手法を用いる [1]。ただし、近年、連続信号を直接モデル化するために、HMM の出力確率分布を連続確率密度関数とした連続型 HMM が用いられることが多い [2]。したがって、本稿においても出力確率分布を単一ガウス分布とした連続型 HMM を用いることとする。以下に、本稿で用いる手法の詳細を述べる。

### 2.1 和音特徴量

和音は、様々な転回形により演奏され、音高配置が異なる場合でも構成音が等しい場合、同一の和音とみなさ

れる。和音の音高配置によらない特徴量として、クロマベクトル [3] がある。クロマベクトルは、パワースペクトルを半音ごとに複数オクターブ間で足し合わせることで得られる。

### 2.2 隠れマルコフモデルを用いた和音認識

和音進行を隠れ状態とし、各時刻で状態から和音特徴量が出力され、和音進行を状態遷移とすることにより、和音認識に HMM を適用できる。HMM を用いた和音認識は、観測された和音特徴量系列からその背後にある和音系列を求める問題となる。観測された和音特徴量系列に対して事後確率が最大となる和音系列は、Viterbi アルゴリズムを用いることにより求められる [1]。

## 3 評価実験

2章で述べた手法に基づく和音認識システムに対し、白色ガウス雑音が付加された学習データと評価データを用いて、認識率の変化について調査した。

### 3.1 実験条件

実験における音響分析条件を表 1、学習データおよび和音モデルの仕様を表 2、評価データの仕様を表 3 に示す。実験用データは、和声理論におけるカデンツ [4] に基づいた和音系列とし、各和音信号は 3 つの正弦波の和として表現することにより用意した。また、実験用データのサンプリング周波数は 44.1kHz、量子化ビット数が 16bit のモノラル信号とした。和音 HMM の学習および認識には、手動で作成した和音名ラベルデータを正解ラベルとして用いた。

### 3.2 実験結果

各 SNR で雑音を付加した学習データを用いて作成した各和音モデルによる認識率を図 1 に示す。横軸が評価データの SNR であり、縦軸が認識率である。図 1 より、学習データの SNR が評価データの SNR と近い場合、高い認識率が得られることを確認できる。例えば、SNR が 10dB の学習データから学習した和音モデルを用いた認識の場合、SNR が 5dB から 15dB の評価データにお

Chord Recognition Using Hidden Markov Model in Noisy Environments

Atsuhiko UEKI<sup>†</sup>, Shunsuke KOSHITA<sup>†</sup>, Masahide ABE<sup>†</sup>, and Masayuki KAWAMATA<sup>†</sup>

<sup>†</sup>Department of Electronic Engineering, Graduate School of Engineering, Tohoku University

表 1: 音響分析条件

|           |              |
|-----------|--------------|
| サンプリング周波数 | 44.1 [kHz]   |
| 分析窓       | 方形窓          |
| 分析窓長      | 0.1 [s]      |
| オーバーラップ長  | 0.0 [s]      |
| 特徴量       | 12 次元クロマベクトル |

表 2: 学習データおよび和音モデルの仕様

|         |                                       |
|---------|---------------------------------------|
| 学習データ   | 和声理論に基づいた和音系列<br>各和音は正弦波の和により表現       |
| 和音の種類   | 12 の音名それぞれにおける<br>major/minor の 24 種類 |
| 重畳雑音    | 平均 0 の白色ガウス雑音                         |
| SNR     | 0, 5, 10, 15, 20 [dB]                 |
| 学習データの数 | 150 曲                                 |
| 和音モデル   | 24 状態 ergodic 連続型 HMM                 |
| 出力確率    | 単一 12 次元ガウス分布                         |

いて高い認識率が得られるが、それ以外の評価データにおいて認識率は大きく減少している。

次に、各和音モデルにおいてどのような評価データに対して最も高い認識率を示すかについて検討するために、図 1 における認識率が 99% から 100% の部分を拡大したものを図 2 に示す。図 2 より、各和音モデルにおいて、それぞれ学習に用いたデータの SNR より、5dB 程度大きい SNR の評価データを対象とした場合に認識率が最大となる傾向があることを確認できる。また、ある SNR の評価データにおいて認識率が最大となったとき、更に 1dB 大きい SNR の評価データに対する認識率は大きく減少することも確認できた。

#### 参考文献

- [1] Sheh and Ellis, “Chord segmentation and recognition using EM-trained hidden markov models,” Proc. 4th ICMIR, pp.183–189, 2003.
- [2] 内山, 宮本, 西本, 小野, 嵯峨山, “調波音・打楽器音分離手法を用いた音楽音響信号からの自動和音認識,” 情報処理学会研究報告, 2008-MUS-76, pp.137–142, 2008.
- [3] Fujishima, “Realtime chord recognition of musical sound: A system using common lisp music,” Proc. ICMC, pp.464–467, 1999.
- [4] 島岡, 野田, 尾高, 川井, 佐藤, 永富, 南, 浦田, 野平, “総合和声 実技・分析・原理,” 音楽之友社, 1998.

表 3: 評価データの仕様

|         |                                       |
|---------|---------------------------------------|
| 評価データ   | 和声理論に基づいた和音系列<br>各和音は正弦波の和により表現       |
| 和音の種類   | 12 の音名それぞれにおける<br>major/minor の 24 種類 |
| 重畳雑音    | 平均 0 の白色ガウス雑音                         |
| SNR     | -10~30 [dB]                           |
| 評価データの数 | 50 曲                                  |

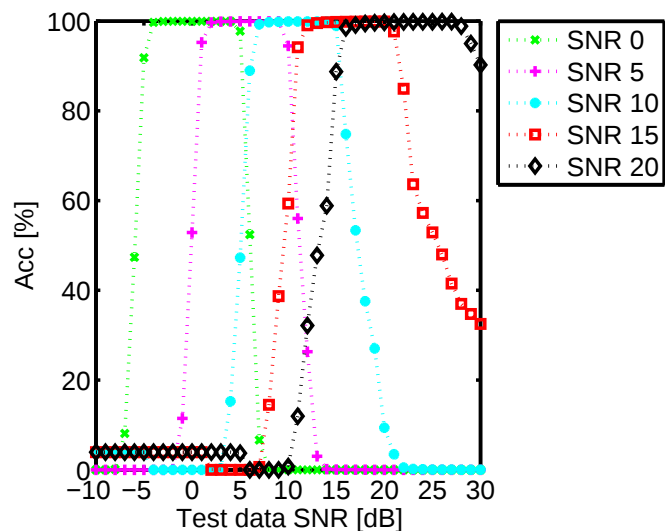


図 1: 評価データの SNR と認識率

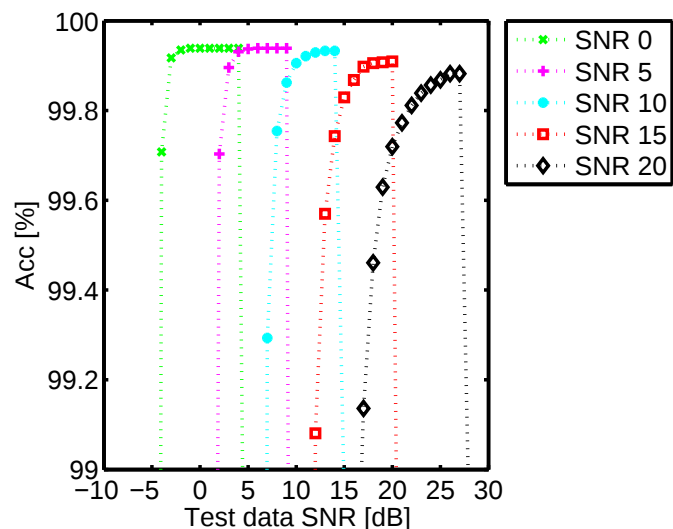


図 2: 評価データの SNR と認識率 (拡大図)